Rates of convergence of Newton type methods for variational inequalities and nonlinear programming

J. Frederic Bonnans

UNITÉ DE RECHERCHE
INRIA-ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France
Tél.:(1) 39 63 55 11

Rapports de Recherche

N° 1260

Programme 5
Automatique, Productique,
Traitement du Signal et des Données

RATES OF CONVERGENCE OF NEWTON TYPE METHODS FOR VARIATIONAL INEQUALITIES AND NONLINEAR PROGRAMMING

Joseph Frédéric BONNANS

Juillet 1990

*RR_1260

Rates of convergence of Newton type methods for variational inequalities and nonlinear programming

Taux de convergence des méthodes de type Newton pour les inéquations variationnelles et la programmation non linéaire

Joseph Frédéric Bonnans*

Abstract

This paper presents some new results in the theory of Newton type methods for variational inequalities, and their application to nonlinear programming. A condition of semi-stability is shown to ensure the quadratic convergence of Newton's method and the superlinear convergence of some quasi-Newton algorithms, provided the sequence defined by the algorithm exists and converges. A partial extension of these results to nonsmooth function is given. The second part of the paper considers some particular variational inequalities with unknowns (x, λ), generalizing optimality systems. Here only the question of superlinear convergence of $\{x^k\}$ is considered. Some necessary or sufficient conditions are given. Applied to some quasi-Newton algorithms they allow to obtain the superlinear convergence of $\{x^k\}$. The application of the previous results to nonlinear programming allows to strenghten the know results, the main point being a characterization of the superlinear convergence of $\{x^k\}$ assuming a weak second-order condition without strict complementarity.

Résumé

Cet article présente quelques résultats nouveaux dans la théorie des méthodes de type Newton pour les inéquations variationnelles, et leur application à la programmation non linéaire. Nous montrons qu'une condition de semi-stabilité implique la convergence quadratique de la méthode de Newton et la convergence superlinéaire de certains algorithmes de quasi-Newton, si la suite définie par l'algorithme existe et converge. Une extension partielle de ces résultats aux fonctions non lisses est présentée. La seconde partie de l'article traité d'inéquations variationnelles particulières à inconnues (x, λ), généralisant les systèmes d'optimalité. Des conditions nécessaires ou suffisantes de convergence superlinéaire de $\{x^k\}$ seul sont exposées. Appliquées à certains algorithmes de quasi-Newton elles permettent d'obtenir la convergence superlinéaire de $\{x^k\}$. L'application des résultats précédents à la programmation non linéaire permet de renforcer les résultats connus, la principale amélioration étant une caractérisation de la convergence superlinéaire de $\{x^k\}$ supposant une condition faible du deuxième ordre, sans complémentarité stricte.

Key Words

Variational inequalities, Linearization, Newton's method, Quasi-Newton algorithms, Successive quadratic programming, Nonsmooth analysis, Stability, Nonlinear programming.

*INRIA, B.P. 105, 78153 Rocquencourt, France

1 Introduction

This paper is devoted to the local study of Newton type algorithms for variational inequalities. Variational inequalities have been studied for a long time (see Lions and Stampacchia [14]) mainly because of their applications to mechanical systems. The operators there being generally monotone, a large theory of monotone operators has been developed (see Brézis [7]) and several algorithms for convex programming, including duality methods, have been extended to this framework (see Gabay [10]). Some problems in economy as well as optimality systems of nonlinear programming problems can also be represented by variational inequalities (see Robinson [18]). Consequently the strenght and large use of Newton type algorithms for nonlinear programming (see Bertsekas [2] and Fletcher [9]) suggests to develop a theory of Newton type methods for variational inequalities.

Some early (but unpublished) works in this direction due to Josephy [12,13] give a local analysis using the concept of strong regularity (Robinson [17]). Josephy obtains a quadratic rate of convergence for Newton's method and superlinear convergence for some quasi-Newton algorithms. In the case of nonlinear programming problems, assuming the gradients of active constraints to be linearly independent, the strong regularity reduces to some strong second-order sufficient condition.

The quadratic rate of convergence under the weak second order sufficiency condition for nonlinear programming problems, and assuming the linear independence of the gradients of active constraints, has been recently obtained by the author [4]. This suggests that the theory of Newton type methods for variational inequalities can be extended. For this purpose we use the new concept of semi-stability. We say that a solution $\bar{x}$ of a variational inequality is semi-stable if, given a small perturbation in the right-hand side, a solution x of the perturbed variational inequality that is sufficiently close to $\bar{x}$, is such that the distance of x to $\bar{x}$ is of the order of the magnitude of the perturbation. Here we do not assume anything about the existence of a solution for the perturbed problem.

Indeed we give a counter example showing that existence for a small perturbation does not always hold under the semi-stability hypothesis. Nevertheless, assuming the existence of the sequence satisfying the Newton steps we show that semi-stability allows to obtain in a simple way quadratic convergence for Newton's method and superlinear convergence for a large class of Newton type algorithms (here we extend the Dennis-Moré's [8] sufficient condition for superlinear convergence). This allows us to adapt Grzegorski's[11] theory in order to derive the superlinear convergence of a large class of quasi-Newton updates including Broyden's one. For polyhedral convex sets we may characterize semi-stability : it reduces to the condition that the solution $\bar{x}$ is an isolated solution of the variational inequality linearized at $\bar{x}$. An equivalent condition is the "strong positivity condition" of Reinoza [16]. We also check that for non-differentiable data the theory can be extended using point-based approximations (reminiscent of those of Robinson [19]) that play the role of linearized function.

The second part of this paper is devoted to a special class of variational inequalities generalizing optimality systems. The unknowns here are couples (x, λ) and we try to obtain conditions related to superlinear convergence of $\{x^k\}$ alone. Indeed we give a characterization of the superlinear convergence of $\{x^k\}$, valid under a second-order hypothesis satisfied by optimality systems for which the weak second-order sufficiency condition holds. This allows us to extend to inequality constrained problems the characterization of Boggs, Tolle and Wang [3] for equality constrained problems (this improves the results of Bonnans [4] in which some necessary or sufficient conditions are given) ; our result assumes only that the gradients of active constraints are linearly independent and the weak second-order sufficient condition holds, but no strict complementarity hypothesis. We apply this result in order to obtain superlinear convergence for a large class of quasi-Newton updates .

Acknowledgement : The author thanks G. Launay from INRIA for her careful reading of a preliminary version of the manuscript, and S.M. Robinson, from the University of Wisconsin-Madison, for his advices.

2 Newton type methods for variational inequalities

Let φ be a continuously differentiable mapping from $\mathbb{R}^q$ into $\mathbb{R}^q$. Given a closed convex subset K of $\mathbb{R}^q$ we consider the variational inequality

$$\langle \varphi(z), y - z \rangle \geq 0, \quad \forall y \in K; \quad z \in K. \tag{2.1}$$

We may define the (closed convex) cone of outward normals to K at a point $z \in K$

$$N(z) := \{x \in \mathbb{R}^q; \quad \langle x, y - z \rangle \leq 0, \quad \forall y \in K\}.$$

A relation equivalent to (2.1) is then

$$\varphi(z) + N(z) \ni 0. \tag{2.2}$$

We define the Newton type algorithm for solving (2.2) as follows :

Algorithm 1
0) Choose $z^0 \in \mathbb{R}^n$, $\quad k \leftarrow 0$.
1) While z^k is not solution of (2.2) : choose M^k, $q \times q$ matrix and compute z^{k+1} solution of

$$\varphi(z^k) + M^k(z^{k+1} - z^k) + N(z^{k+1}) \ni 0. \tag{2.3}$$

$\square$

In order to obtain estimates of the rate of convergence of $\{z^k\}$ we essentially use the following concept.

Definition 2.1 *A solution $\bar{z}$ of (2.2) is said to be semi-stable if there exist $c_1 > 0$ and $c_2 > 0$ such that, for all $(z, \delta) \in \mathbb{R}^q \times \mathbb{R}^q$ solution of*

$$\varphi(z) + N(z) \ni \delta,$$

and $\|z - \bar{z}\| \leq c_1$, then $\|z - \bar{z}\| \leq c_2 \|\delta\|$.

Theorem 2.1 *Let $\bar{z}$ be a semi-stable solution of (2.1), and let $\{z^k\}$ computed by Algorithm 1 converge towards $\bar{z}$. Then* *(i) If* $(\varphi'(\bar z) - M^k)(z^{k+1} - z^k) = o(z^{k+1} - z^k)$ *then* $\{z^k\}$ *converges superlinearly.*

(ii) If $(\varphi'(\bar z) - M^k)(z^{k+1} - z^k) = 0(\|z^{k+1} - z^k\|^2)$ *and* φ' *is locally Lipschitz then* $\{z^k\}$ *converges quadratically.*

Proof Define $\theta^k := (\varphi'(\bar z) - M^k)(z^{k+1} - z^k)$. We can write the Newton type step (2.3) as

$$\varphi(z^k) + \varphi'(\bar z)(z^{k+1} - z^k) + N(z^{k+1}) \ni \theta^k,$$

hence

$$\varphi(z^{k+1}) + N(z^{k+1}) \ni r^k \tag{2.4}$$

with, using the continuity of φ' at $\bar z$:

$$r^k := \theta^k + \varphi(z^{k+1}) - \varphi(z^k) - \varphi'(\bar z)(z^{k+1} - z^k) = \theta^k + o(z^{k+1} - z^k).$$

If $\theta^k = o(z^{k+1} - z^k)$ then from the semi-stability of $\bar z$ and (2.4) we get

$$z^{k+1} - \bar z = 0(r^k) = o(z^{k+1} - z^k) = o(\|z^{k+1} - \bar z\| + \|z^k - \bar z\|)$$

hence $z^{k+1} - \bar z = o(z^k - \bar z)$, i.e. $\{z^k\}$ converges superlinearly. This proves (i). If φ' is locally Lipschitz and $\theta^k = 0(\|z^{k+1} - z^k\|^2)$ we already know that $\{z^k\}$ converges superlinearly, hence $\|z^{k+1} - z^k\|/\|z^k - \bar z\| \to 1$. Let L be a Lipschitz constant of φ' at $\bar z$. This implies, for k large enough :

$$\begin{aligned} \|\varphi(z^{k+1}) - \varphi(z^k) - \varphi'(\bar z)(z^{k+1} - z^k)\| &= \|\int_0^1 \left[\varphi'(z^k + \sigma(z^{k+1} - z^k)) - \varphi'(\bar z)\right](z^{k+1} - z^k)d\sigma\| \\ &\leq L\max(\|z^{k+1} - \bar z\|, \|z^k - \bar z\|)\|z^{k+1} - z^k\| \\ &\leq 2L\|z^{k+1} - z^k\|^2, \end{aligned}$$

hence,

$$z^{k+1} - \bar z = 0(r^k) = 0(\|z^{k+1} - z^k\|^2) = 0(\|z^k - \bar z\|^2),$$

from which the quadratic convergence follows. □

Corollary 2.1 *If* $\{z^k\}$ *computed by Algorithm 1 converges toward a semi-stable solution* $\bar z$ *of (2.1) then*

(i) If $M^k \to \varphi'(\bar z)$, $\{z^k\}$ *converges superlinearly.*

(ii) If φ' *is locally Lipschitz and* $M^k = \varphi'(\bar x) + 0(z^k - \bar z)$ *(which is the case for Newton's method under the hypothesis of Lipschitz continuity of* φ'*) then* $\{z^k\}$ *converges quadratically.*

Until now we assumed the existence of a converging sequence instead of giving the hypotheses that imply its existence. Our point of view is that it is clearer to do so ; indeed, if we want now to prove that the sequence is well defined for say Newton's method with a good starting point, we just have to posit the following definition :

Definition 2.2 *We will say that* $\bar z$ *is a hemi-stable solution of (2.1) if for all* $\alpha > 0$ *there exists* $\varepsilon > 0$ *such that, given* $\hat z \in \mathbb{R}^q$, *the variational inequality (in z)*

$$\varphi(\hat z) + M(z - \hat z) + N(z) \ni 0$$

has a solution z satisfying $\|z - \bar z\| \leq \alpha$, *whenever* $\|\hat z - \bar z\| + \|M - \varphi'(\bar z)\| < \varepsilon$.

Then from Corollary 2.1, we easily obtain

Theorem 2.2 *If $\bar{z}$ is a semi-stable and hemi-stable solution of (2.1), there exists $\varepsilon > 0$ such that if $\|z^o - \bar{z}\| \leq \varepsilon$, and assuming $\{z^k\}$ is chosen so that $\|z^k - \bar{z}\| \leq \varepsilon$:*

(i) *At each step k there exists z^{k+1} solution of the Newton step satisfying $\|z^{k+1} - \bar{z}\| \leq \varepsilon$.*

(ii) *The sequence $\{z^k\}$ defined in this way converges superlinearly (quadratically if φ' is locally Lipschitz) towards $\bar{z}$.*

□

Remark 2.1 **(i)** *Semi-stability does not imply hemi-stability, as shown by the following example : consider for $\varepsilon \in \mathbb{R}$ the variational inequality*

$$(\varepsilon - z)(y - z) \geq 0, \ \forall y \geq 0; \ z \geq 0,$$

corresponding to the badly posed optimization problem

$$\min \varepsilon z - z^2/2; \ z \geq 0.$$

For $\varepsilon = 0$, $\bar{z} = 0$ is the unique solution. As $\varphi(z) = \varepsilon - z$ is linear, semi-stability holds. However when $\varepsilon < 0$ the variational inequality has no solution.

(ii) *A sufficient condition for semi and hemi-stability is the strong regularity of Robinson [17].*

(iii) *We will see later that in the case of optimality systems for local solutions of nonlinear programming problems, semi-stability and hemi-stability are equivalent.*

Theorem 2.1 may also be used in order to derive superlinear convergence of some quasi-Newton algorithm. By quasi-Newton algorithm we mean a Newton type algorithm with M^k satisfying the so-called quasi-Newton equation

$$M(z^{k+1} - z^k) = \varphi(z^{k+1}) - \varphi(z^k). \tag{2.5}$$

A typical situation is when a closed convex subset $\mathcal{K}$ of the space of $q \times q$ matrices is known to satisfy

$$\varphi'(z) \in \mathcal{K}, \ \forall z \in \mathbb{R}^q. \tag{2.6}$$

Then M^{k+1} is taken as a solution of

$$\min \|M - M^k\|_{\sharp} ; \ M \in \mathcal{K} \quad \text{and} \quad M \text{ satisfies (2.5)}. \tag{2.7}$$

Here $\|\cdot\|_{\sharp}$ is a matrix norm that we will assume to be associated to a scalar product. When $\|.\|_{\sharp}$ is the Frobenius norm we recover Broyden's update when $\mathcal{K}$ is the space of $q \times q$ matrices, the PSB update when $\mathcal{K}$ is the space of symmetric matrices, etc.; see Grzegòrski [11]). We first quote

Lemma 2.1 *Under the hypotheses of Theorem 2.1, if $\{M^k\}$ satisfies the quasi-Newton equation and $(M^{k+1} - M^k)(z^{k+1} - z^k) = o(z^{k+1} - z^k)$, then $\{z^k\}$ converges superlinearly.*

Proof Using (2.5) we get

$$\begin{aligned}(M^{k+1}-M^k)(z^{k+1}-z^k) &= \varphi(z^{k+1})-\varphi(z^k)-M^k(z^{k+1}-z^k)\\ &= (\varphi'(\bar z)-M^k)(z^{k+1}-z^k)+o(z^{k+1}-z^k).\end{aligned}$$

The conclusion is then obtained with Theorem 2.1. □

Theorem 2.3 *Let $\bar z$ be a semi-stable solution of (2.2) and let $\{z^k\}$ be computed by Algorithm 1 converge towards $\bar z$. We assume that (2.5)-(2.7) hold and that*

$$\sum_k \|z^k-\bar z\| < \infty. \tag{2.8}$$

Then $\{z^k\}$ converges superlinearly.

Proof : Define

$$S^k := \{M \in \mathcal{K};\ M(z^{k+1}-z^k) = \varphi(z^{k+1})-\varphi(z^k)\}.$$

Then M^{k+1} is the projection of M^k onto S^k (with the $\|\cdot\|_\sharp$ norm), hence for all $M \in S^k$ we have (see Grzegòrski [11])

$$\|M^{k+1}-M^k\|_\sharp^2 + \|M^{k+1}-M\|_\sharp^2 \le \|M^k-M\|_\sharp^2, \tag{2.9}$$

and a fortiori

$$\|M^{k+1}-M\|_\sharp \le \|M^k-M\|_\sharp. \tag{2.10}$$

Define

$$\begin{aligned}\psi^k &:= \int_0^1 \varphi'(z^k+\sigma(z^{k+1}-z^k))d\sigma,\\ \nu^k &:= \max(\|z^{k+1}-\bar z\|, \|z^k-\bar z\|).\end{aligned}$$

Then ψ^k is an element of S^k and, for k large enough we have, L being a Lipschitz constant of φ' at $\bar z$ in the $\|\cdot\|_\sharp$ norm :

$$\|\psi^k-\varphi'(\bar z)\|_\sharp = \|\int_0^1 \left[\varphi'(z^k+\sigma(z^{k+1}-z^k))-\varphi'(\bar z)\right] d\sigma\|_\sharp \le L\nu^k,$$

hence taking $M = \psi^k$ in (2.10) and using the previous inequality we get

$$\|M^{k+1}-\varphi'(\bar z)\|_\sharp \le \|M^k-\varphi'(\bar z)\|_\sharp + 2L\nu^k. \tag{2.11}$$

This bounded deterioration result and (2.8) imply that $\|M^k-\varphi(\bar z)\|_\sharp$ converges (see Dennis-Moré [8]). As $\psi^k \to \varphi'(\bar z), \|M^k-\psi^k\|_\sharp$ and $\|M^{k+1}-\psi^k\|_\sharp$ also converge towards the same limit. Taking $M = \psi^k$ in (2.9) we deduce that $M^{k+1}-M^k \to 0$; this and Lemma 2.1 imply the conclusion. □

Remark 2.2 When $K = \mathbb{R}^q$, i.e. (2.2) is an equation, if we assume that (z^o, M^o) is close to $(\bar z, \varphi'(\bar z))$ then (assuming $\varphi'(\bar z)$ invertible) the sequence $\{z^k\}$ is well defined and converges superlinearly towards $\bar z$. This type of result can be easily extended. In fact using Theorem 2.2, it is sufficient to prove that $\{z^k\}$ converges linearly and here the key point is again the bounded deterioration (2.11). An hypothesis like hemi-stability is, however, required in order to obtain the existence of the sequence $\{z^k\}$.

3 Characterization of semi-stability when K is polyhedral

We assume here that K is polyhedral, i.e. defined by a finite number of linear equalities and inequalities. This allows us to give several characterizations of semi-stability.

Theorem 3.1 *If K is polyhedral and $\bar{z}$ is a solution of (2.1), $\bar{z}$ is semi-stable iff one of the following hypotheses holds :*

(a) $\bar{z}$ is an isolated solution of the linearization at $\bar{z}$ of (2.2) :

$$\varphi(\bar{z}) + \varphi'(\bar{z})(z - \bar{z}) + N(z) \ni 0. \tag{3.1}$$

(b) One has $\langle z - \bar{z}, \varphi'(\bar{z})(z - \bar{z})\rangle > 0$ for all z different of $\bar{z}$ solution of

$$\begin{aligned} &\langle \varphi(\bar{z}), z - \bar{z}\rangle = 0, && (i)\\ &\varphi(\bar{z}) + \varphi'(\bar{z})(z - \bar{z}) + N(\bar{z}) \ni 0. && (ii) \end{aligned} \tag{3.2}$$

(c) The relation below has no solution but $\bar{z}$:

$$\begin{aligned} &N(z) \subset N(\bar{z}), && (i)\\ &\langle \varphi(\bar{z}), z - \bar{z}\rangle = 0, && (ii)\\ &\alpha\varphi(\bar{z}) + \varphi'(\bar{z})(z - \bar{z}) + N(z) \ni 0, \text{ for some } \alpha \geq 0. && (iii) \end{aligned} \tag{3.3}$$

Remark 3.1 *Theorem 3.1 is connected to the results of Reinoza [16] in the following way. Reinoza states condition (b) and calls it a strong positivity condition (although in the context of non linear programming we will see that it corresponds to weak second-order sufficient conditions, hence it might be better to call it a weak positivity condition). Then Reinoza states that condition (b) is equivalent to condition (d) below :*

$$\begin{cases} \bar{z} \quad \text{is an isolated solution of} \\ \varphi(\bar{z}) + \varphi'(\bar{z})(z - \bar{z}) + N(\bar{z}) \ni 0. \end{cases}$$

The above condition differs from condition (a) by the term $N(\bar{z})$ instead of $N(z)$. Now as we prove that condition (a) and (b) are equivalent, it follows that if (a) is not equivalent to (d), the assertion of Reinoza is false. Take the simple example

$$z(y - z) \geq 0;\ \forall y \geq 0;\ z \geq 0.$$

Here $\bar{z} = 0$ satisfies (a) but as $N(\bar{z}) = \mathbf{R}^-$, any $z \geq 0$ is solution of $z + N(\bar{z}) \ni 0$. Hence (a) is not equivalent to (d).

Proof of Theorem 3.1 : We will prove that

$$\{\bar{z} \text{ is semi-stable}\} \Rightarrow \text{(a)} \Rightarrow \text{(b)} \Rightarrow \text{(c)} \Rightarrow \{\bar{z} \text{ is semi-stable}\}.$$

a) Proof of $\{\bar{z}$ semi-stable$\} \Rightarrow$ (a). If z is solution of (3.1) then from the first order expansion of φ at $\bar{z}$:

$$\varphi(z) + N(z) \ni o(z - \bar{z}),$$

hence if $\bar{z}$ is semi-stable and $\|z-\bar{z}\| \le c_1$, we get $\|z-\bar{z}\| = o(z-\bar{z})$ and this implies $z=\bar{z}$ for z close enough to $\bar{z}$; hence (a) holds.

b) Proof of (a) $\Rightarrow$ (b). Let z contradict (b), i.e. $z \ne \bar{z}$, z satisfies (3.2) but $\langle z-\bar{z}, \varphi'(\bar{z})(z-\bar{z})\rangle \le 0$. From (3.2) we get

$$0 \le \langle \varphi(\bar{z}) + \varphi'(\bar{z})(z-\bar{z}), z-\bar{z}\rangle = \langle z-\bar{z}, \varphi'(\bar{z})(z-\bar{z})\rangle$$

hence

$$\langle z-\bar{z}, \varphi'(\bar{z})(z-\bar{z})\rangle = 0. \tag{3.4}$$

For α in $]0,1[$ define $z^\alpha := \bar{z} + \alpha(z-\bar{z})$. From (3.2 ii), (2.2) and the convexity of $N(\bar{z})$ we deduce that

$$\varphi(\bar{z}) + \varphi'(\bar{z})(z^\alpha - \bar{z}) + N(\bar{z}) \ni 0,$$

hence with (3.2 i) and (3.4), for all $y \in K$:

$$\begin{aligned} 0 &\le \langle \varphi(\bar{z}) + \varphi'(\bar{z})(z^\alpha - \bar{z}), y-\bar{z}\rangle, \\ &= \langle \varphi(\bar{z}) + \varphi'(\bar{z})(z^\alpha - \bar{z}), y-z^\alpha\rangle, \end{aligned}$$

that is

$$\varphi(\bar{z}) + \varphi'(\bar{z})(z^\alpha - \bar{z}) + N(z^\alpha) \ni 0,$$

hence z^α is a solution of (3.1). Also $z^\alpha \to \bar{z}$ when $\alpha \searrow 0$; this contradicts (a).

c) Proof of (b) $\Rightarrow$ (c). Assume that (c) does not hold and let $z \ne \bar{z}$ be a solution of (3.3). From (3.3 ii-iii) we deduce that

$$\langle z-\bar{z}, \varphi'(\bar{z})(z-\bar{z})\rangle \le 0.$$

As (3.2 i) coincide with (3.3 ii) it remains to derive (3.2 ii) in order to get a contradiction with (ii). If $\alpha \le 1$, multiplying relation (2.2) by $(1-\alpha)$, adding it to (3.3 iii) and using (3.3 i) we get (3.2 ii). If $\alpha > 1$ we may check similarly, dividing (3.3 iii) by α, that $y^\alpha := \bar{z} + \frac{1}{\alpha}(z-\bar{z})$ contradicts (b).

d) Proof of (c) $\Rightarrow$ $\{\bar{z}$ is semi-stable$\}$. If $\bar{z}$ is not semi-stable let $z^k \to \bar{z}$ and $\delta^k \to 0$ in $\mathbb{R}^n$ be such that

$$\varphi(z^k) + N(z^k) \ni \delta^k \tag{3.5}$$

and $\|\delta^k\|/\|z^k - \bar{z}\| \to 0$. Define $\beta^k := \|z^k - \bar{z}\|^{-1}$ and $w^k := \beta^k(z^k - \bar{z})$. Then substituting $\varphi(\bar{z}) + \varphi'(\bar{z})(z^k - \bar{z}) + o(z^k - \bar{z})$ to $\varphi(z^k)$ in (3.5) we get after multiplication by β^k

$$\beta^k\varphi(\bar{z}) + \varphi'(\bar{z})w^k + N(z^k) \ni \beta^k\delta^k + \beta^k o(z^k - \bar{z}). \tag{3.6}$$

This right-hand side of (3.6) has limit 0. As K is a polyhedron we may extract without loss of generality a subsequence such that $N(z^o) = N(z^k)$ for all k; also $\|w^k\| = 1$ hence $\{w^k\}$ has at least a limit-point w with $\|w\| = 1$. Again as K is a polyhedron, the set $N^0 := N(z^o) + \mathbb{R}^+\varphi(\bar{z})$ is the cone of exterior normals at z^o to the set

$$K^0 := K \cap \{z \in \mathbb{R}^q; \langle z - z^o, \varphi(\bar{z})\rangle \le 0\}.$$

Hence N^0 is closed. By (3.6) and the closedness of N^0

$$\mathbb{R}^+\varphi(\bar{z}) + \varphi'(\bar{z})w + N(z^o) \ni 0. \tag{3.7}$$

Also as $\beta^k \geq 0$ and the vectors $\bar{z} + (\beta^k)^{-1}w^k = z^k$, $z^k - (\beta^k)^{-1}w^k = \bar{z}$ are elements of K we get from (2.1) and (3.5)

$$\begin{cases} \langle w^k, \varphi(\bar{z})\rangle & = & \beta^k\langle z^k - \bar{z}, \varphi(\bar{z})\rangle \geq 0, \\ -\langle w^k, \varphi(z^k)\rangle & = & \beta^k\langle \bar{z} - z^k, \varphi(z^k)\rangle \geq \beta^k\langle \bar{z} - z^k, \delta^k\rangle \to 0. \end{cases} \tag{3.8}$$

As $\delta^k \to 0$ and $z^k \to \bar{z}$, $\varphi(z^k) \to \varphi(\bar{z})$. This, (3.8) and $w^k \to w$ imply

$$\langle w, \varphi(\bar{z})\rangle = 0. \tag{3.9}$$

Now, as K is a polyhedron, $\bar{z} + \varepsilon w$ is in K for $\varepsilon > 0$ small enough. Let us check that $N(\bar{z} + \varepsilon w) \supset N(z^o)$. It is sufficient to check that any linear inequality constraint defining K that is active at z^o is also active at $\bar{z} + \varepsilon w$. Here we say that a constraint $\langle a, z\rangle \leq b$ is active at z if $\langle a, z\rangle = b$. Extracting again if necessary a subsequence we may assume that the set of active constraints is the same for all $\{z^k\}$. Then for the subsequence considered here we have $\langle a, z^k\rangle = b$, hence $\langle a, \bar{z}\rangle = b$ and $\langle a, w^k\rangle = 0$, from which $\langle a, w\rangle = 0$, and finally $\langle a, \bar{z} + \varepsilon w\rangle = b$. This proves that $N(\bar{z} + \varepsilon w) \supset N(z^o)$. This and (3.7) (multiplied by $\varepsilon > 0$) imply

$$\mathbb{R}^+\varphi(\bar{z}) + \varepsilon\varphi'(\bar{z})w + N(\bar{z} + \varepsilon w) \ni 0. \tag{3.10}$$

Also for $\varepsilon > 0$ small enough and as K is a polyhedron, $N(\bar{z} + \varepsilon w) \subset N(\bar{z})$.This, (3.9), (3.10) and the fact that $z = \bar{z} + \varepsilon w$ is in K give a contradiction to (c). □

Remark 3.2 *The proof of $\{\bar{z}$ is semi-stable$\}$ ⇒ (a) ⇒ (b) ⇒ (c) does not use the fact that K is polyhedral, and is still valid in a Hilbert space.*

4 Extension of the theory to nonsmooth data

Although we are mainly interested in this paper by finite dimensional variational inequalities with smooth data we will give here an extension of the previous results to problems in a Hilbert space with nonsmooth data. Let K be a closed convex subset of a Hilbert space Z, $N(z)$ the cone of outward normals to K at z and φ a mapping from Z into itself. In order to define an extension of Algorithm 1 for the problem

$$\varphi(z) + N(z) \ni 0, \tag{4.1}$$

we use a concept of point-based approximation (PBA) close to the one of Robinson [19].

Definition 4.1 *We say that $\psi : Z \times Z \to Z$ is a PBA to φ if for any two sequences $\{y^k\}, \{z^k\}$ converging to the same point the following holds :*

$$\|\varphi(y^k) - \psi(z^k, y^k)\| \leq r(y^k, z^k) \tag{4.2}$$

with $r(y^k, z^k)/\|y^k - z^k\| \to 0$.

Here $\psi(z^k,.)$ can be seen as a generalization of the linearization of φ at z^k (see Remark 4.1 below). We now define a somewhat abstract Newton type method as the following algorithm :

Algorithm 2
0) Choose $z^0 \in X; k \leftarrow 0$.

1) While z^k does not satisfy (4.1) : choose a mapping $\Xi^k : X \to X$, approximation of $\psi(z^k,.)$. Compute z^{k+1} solution of

$$\Xi^k(z^{k+1}) + N(z^{k+1}) \ni 0. \qquad \square \tag{4.3}$$

We define semi-stability as in section 2.

Theorem 4.1 *If $\{z^k\}$ computed by Algorithm 2 converges towards a semi-stable solution $\bar{z}$ of (4.1), then*

(i) If $\psi(z^k, z^{k+1}) - \Xi^k(z^{k+1}) = o(z^{k+1} - z^k)$, then $\{z^k\}$ converges superlinearly.

(ii) If $\psi(z^k, z^{k+1}) - \Xi^k(z^{k+1}) = 0(\|z^{k+1} - z^k\|^2)$ and for some $c_1 > 0$ and all (y, z) close enough to $\bar{z}$ the function r in (4.2) satisfies $r(y, z) \le c_1\|y - z\|^2$ then $\{x^k\}$ converges quadratically.

Proof Writing the step (4.3) as

$$\psi(z^k, z^{k+1}) + N(z^{k+1}) \ni \psi(z^k, z^{k+1}) - \Xi^k(z^{k+1})$$

and using (4.2), we deduce that

$$\varphi(z^{k+1}) + N(z^{k+1}) \ni \psi(z^k, z^{k+1}) - \Xi^k(z^{k+1}) + o(z^{k+1} - z^k).$$

In case (i) it follows from semi-stability that $z^{k+1} - \bar{z} = o(z^{k+1} - z^k)$, hence z^k converges superlinearly. In case (ii) we similarly obtain $z^{k+1} - \bar{z} = 0(\|z^{k+1} - z^k\|^2)$,which implies the quadratic convergence. $\square$

Remark 4.1 *Theorem 4.1 can be seen as an extension of Theorem 2.1. Indeed if φ is continuously differentiable and*

$$\psi(z^k, z^{k+1}) = \varphi(z^k) + \varphi'(z^k)(z^{k+1} - z^k),$$

$$\Xi^k(z^{k+1}) = \varphi(z^k) + M^k(z^{k+1} - z^k),$$

for some $q \times q$ matrix M^k, then

$$\begin{aligned} \psi(z^k, z^{k+1}) - \Xi^k(z^{k+1}) &= (\varphi'(z^k) - M^k)(z^{k+1} - z^k) \\ &= (\varphi'(\bar{z}) - M^k)(z^{k+1} - z^k) + o(z^{k+1} - z^k), \end{aligned}$$

hence point (i) of Theorem 4.1 reduces to point (i) of Theorem 2.1. Similarly if φ' is locally Lipschitz we have

$$(\varphi'(z^k) - M^k)(z^{k+1} - z^k) = (\varphi'(\bar{z}) - M^k)(z^{k+1} - z^k) + 0(\|z^k - \bar{z}\|.\|z^{k+1} - z^k\|),$$

the last term being $0(\|z^{k+1} - z^k\|)^2$ as $\|z^{k+1} - z^k\|/\|z^k - \bar{z}\| \to 1$ because of the superlinear convergence, hence point (ii) of Theorem 4.1 reduces to point (ii) of Theorem 2.1.

We define the directional derivatives $\varphi'(.,.)$ of φ as the limit

$$\varphi'(z,d) := \lim_{\alpha \searrow 0} \frac{1}{\alpha}[\varphi(z+\alpha d) - \varphi(z)].$$

We will state in Theorem 4.2 below an extension of Theorem 3.1. This Theorem 4.2 applies to mappings having the following property : φ has directional derivatives and for all z and any sequence $\{y^k\}$ converging towards z

$$\varphi(y^k) = \varphi(z) + \varphi'(z, y^k - z) + o(y^k - z). \tag{4.4}$$

We give an example of mappings satisfying (4.4).

Lemma 4.1 *Assume that $X = \mathbb{R}^q$ and φ is a Lipschitz mapping having continuous directional derivatives (i.e. for all $z \in \mathbb{R}^q$, the mapping $d \rightarrow \varphi'(z,d)$ is continuous). Then (4.4) is satisfied.*

Proof If the conclusion is false, there exists $\{y^k\}$ converging towards z and contradicting (4.4). Extracting a subsequence if necessary we may assume that $w^k := (y^k - z)/\|y^k - z\|$ has a limit w and that $\liminf \|\varphi(y^k) - \varphi(z) - \varphi'(z, y^k - z)\|/\|y^k - z\| > 0$. But as φ is Lipschitz

$$\begin{aligned}
\varphi(y^k) &= \varphi(z + \|y^k - z\|w + \|y^k - z\|(w^k - w)), \\
&= \varphi(z + \|y^k - z\|w) + o(y^k - z), \\
&= \varphi(z) + \|y^k - z\|\varphi'(z,w) + o(y^k - z), \\
&= \varphi(z) + \varphi'(z, y^k - z) + \|y^k - z\|(\varphi'(z,w) - \varphi'(z,w^k)) + o(y^k - z).
\end{aligned}$$

Then the continuity of the directional derivatives of φ implies that (4.4) is satisfied, in contradiction with our hypothesis. $\square$

The proof of Theorem 4.2 below being a direct extension of the one of Theorem 3.1 we do not give it (let us say however that (4.4) is used to extend (3.6)).

Theorem 4.2 *Assume that $X = \mathbb{R}^q$, φ is a Lipschitz mapping with continuous directional derivatives, K is polyhedral and $\bar{z}$ is a solution of (4.1). Then $\bar{z}$ is semi-stable iff one of the following hypotheses holds :*

(a) $\bar{z}$ is an isolated solution of the linearization at $\bar{z}$ of (4.1) defined as follows :

$$\varphi(\bar{z}) + \varphi'(\bar{z}, z - \bar{z}) + N(z) \ni 0.$$

(b) One has $\langle z - \bar{z}, \varphi'(\bar{z}, z - \bar{z})\rangle > 0$ for all z different of $\bar{z}$ solution of

$$\langle \varphi(\bar{z}), z - \bar{z}\rangle = 0,$$

$$\varphi(\bar{z}) + \varphi'(\bar{z}, z - \bar{z}) + N(\bar{z}) \ni 0.$$

(c) The relation below has no solution but $\bar{z}$:

$$N(z) \subset N(\bar{z})$$

$$\langle \varphi(\bar{z}), z - \bar{z}\rangle = 0$$

$$\alpha\varphi(\bar{z}) + \varphi'(\bar{z}, z - \bar{z}) + N(z) \ni 0 \quad \textit{for some } \alpha \geq 0.$$

5 Convergence analysis for some structured variational inequalities

We now specialize our study to a particular case of variational inequalities. In the next section we will apply the results of this section to nonlinear programming problems. Let F, g be smooth mappings : $\mathbb{R}^n \to \mathbb{R}^n$ and $\mathbb{R}^n \to \mathbb{R}^p$, respectively. Let I, J be a partition of $\{1, \ldots, p\}$. By $g(x) \ll 0$ we mean

$$\begin{aligned} g_i(x) &\leq 0, \forall i \in I, \\ g_j(x) &= 0, \forall j \in J. \end{aligned}$$

We now consider the system (in which $\lambda \in \mathbb{R}^p$)

$$\left\{ \begin{array}{l} F(x) + g'(x)^* \lambda = 0, \\ \\ g(x) \ll 0, \ \lambda_I \geq 0, \ \lambda_i g_i(x) = 0, \ \forall i \in I \cup J. \end{array} \right. \tag{5.1}$$

As observed in Robinson [19] we may embed (5.1) into (2.1) in the following way. Put $q := n + p$, $z := (x, \lambda)$ and

$$\varphi(x, \lambda) := \begin{pmatrix} F(x) + g'(x)^* \lambda \\ -g(x) \end{pmatrix};$$

$$K_1 := \{\lambda \in \mathbb{R}^p, \lambda_I \geq 0\} \ ; \ K := \mathbb{R}^n \times K_1;$$

so that K is polyhedral and

$$N(x, \lambda) = \{0\} \times N_1(\lambda),$$

with $N_1(\lambda)$ normal cone (o cone of outwards normals) to K_1 at λ, i.e.

$$N_1(\lambda) = \{\mu \in \mathbb{R}^p; \ \mu_J = 0; \ \mu_I \leq 0; \ \mu_i = 0 \text{ if } \lambda_i > 0, \forall i \in I\}.$$

The corresponding variational inequality can be written in the following way :

$$\left\{ \begin{array}{lll} F(x) & +g'(x)^* \lambda & = 0, \\ -g(x) & +N_1(\lambda) & \ni 0. \end{array} \right. \tag{5.2}$$

Let us denote

$$H(x, \lambda) := F'(x) + \sum_{i=1}^{p} \lambda_i \nabla^2 g_i(x).$$

Then we have

$$\varphi'(x, \lambda) = \begin{pmatrix} H(x, \lambda) & g'(x)^* \\ \\ -g'(x) & 0 \end{pmatrix}, \tag{5.3}$$

and

$$\langle (y, \mu), \varphi'(x, \lambda)(y, \mu) \rangle = \langle y, H(x, \lambda) y \rangle. \tag{5.4}$$

Taking (5.2)-(5.3) and Theorem 3.1 in account, we see that semi-stability for (5.2) (expressed at some point $(\bar{x}, \bar{\lambda})$ solution of (5.2)) can be stated as

$$
\begin{cases}
(y, \mu) = 0 \text{ is an isolated solution of} \\
(i) \quad H(\bar{x}, \bar{\lambda})y + g'(\bar{x})^* \mu = 0, \\
(ii) \quad g(\bar{x}) + g'(\bar{x})y \in N_1(\bar{\lambda} + \mu).
\end{cases}
\tag{5.5}
$$

For any $\hat{I} \subset I$ by $z \overset{\hat{I}}{\ll} 0$ we mean $z_J = 0$ and $z_i \leq 0$ for all i in $\hat{I}$. Let us define

$$
\begin{aligned}
\bar{I} &:= \{i \in I; g_i(\bar{x}) = 0\}, \\
I^+ &:= \{i \in \bar{I}; \bar{\lambda}_i > 0\}, \\
I^0 &:= \bar{I} - I^+ = \{i \in \bar{I};\ \bar{\lambda}_i = 0\}, \\
I^* &:= J \cup I^+.
\end{aligned}
$$

It may be convenient to define the so-called "critical cone" (or cone of critical directions) :

$$
C = \{y \in \mathbf{R}^n;\ g'(\bar{x})y \overset{\bar{I}}{\ll} 0;\ g'_{I^+}(\bar{x})y = 0\}.
$$

Proposition 5.1 *Hypothesis (5.5) is equivalent to*

$$
\begin{cases}
(y, \mu) = 0 \textit{ is the unique solution of} \\
(i) \quad H(\bar{x}, \bar{\lambda})y + g'(\bar{x})^* \mu = 0, \\
(ii) \quad y \in C,\ \mu_{I^0} \geq 0;\ \mu_i = 0 \textit{ if } g_i(\bar{x}) < 0, \forall i \in I;\ \mu_i g'_i(\bar{x})y = 0,\ \forall i \in I^0.
\end{cases}
\tag{5.6}
$$

Proof The set of solutions of (5.6.i-ii) is a cone. Hence it is equivalent to state that $(y, \mu) = 0$ is the unique solution of (5.6i-ii) or to state that $(y, \mu) = 0$ is an isolated solution of (5.6i-ii)). Now it is sufficient to prove the equivalence of (5.5 ii) and (5.6 ii) when (y, μ) is small enough. If μ is close to zero then $\bar{\lambda}_i + \mu_i = 0$ only if $\bar{\lambda}_i = 0$. On the other hand μ_{I^0} if (5.5ii) holds must be nonnegative. For that reason (5.5 ii) is equivalent (when (y, μ) is small enough) to

$$
\begin{cases}
g'_i(\bar{x})y &= 0, \quad \forall i \in J \cup I^+, \\
g'_i(\bar{x})y &\leq 0,\ \mu_i \geq 0, \mu_i g'_i(\bar{x})y = 0,\ \forall i \in I^0, \\
\mu_i &= 0 \quad \text{if } g_i(\bar{x}) < 0,
\end{cases}
$$

and this is easily shown to be equivalent to (5.6 ii). □

Let us now consider Newton's method applied to (5.2). The subproblem to be solved at step k is, denoting by d^k the increment in x :

$$
\begin{cases}
F(x^k) \quad + H(x^k, \lambda^k)d^k + g'(x^k)^* \lambda^{k+1} = 0, \\
g(x^k) \quad + g'(x^k)d^k \in N_1(\lambda^{k+1}).
\end{cases}
$$

As the evaluation of $g'(x^k)$ is already necessary in order to evaluate $\varphi(x^k, \lambda^k)$ the only part of the Jacobian that perhaps needs to be approximated is $H(x^k, \lambda^k)$. We obtain then the Newton type algorithm

Algorithm 3
0) Choose $(x^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^p$. $k \leftarrow 0$.
1) While (x^k, λ^k) is not solution of (5.2) : Choose $M^k, n \times n$ matrix and compute (d^k, λ^{k+1}) solution of

$$\begin{cases} F(x^k) & +M^k d^k + g'(x^k)^* \lambda^{k+1} = 0, \\ g(x^k) & +g'(x^k) d^k \in N_1(\lambda^{k+1}). \end{cases}$$

□

When $M^k = H(x^k, \lambda^k)$, applying Corollary 2.1 and Proposition 5.1, we easily obtain

Theorem 5.1 *(Convergence of Newton's method). Let $\{x^k, \lambda^k\}$ be computed by Algorithm 3 with $M^k = H(x^k, \lambda^k)$ converge toward $(\bar{x}, \bar{\lambda})$ satisfying (5.2) and (5.6). If $x \rightarrow (F(x), g'(x))$ is C^1 (resp. locally Lipschitz) then $(x^k, \lambda^k) \rightarrow (\bar{x}, \bar{\lambda})$ superlinearly (resp. at a quadratic rate).*

We now consider conditions related to the superlinear convergence of $\{x^k\}$ alone. We are looking for necessary and/or sufficient conditions of the following type : at each iteration k we define

$$\begin{array}{ll} E^k & \text{closed convex subset of } \mathbb{R}^n, \\ P^k & \text{orthogonal projection onto } E^k, \\ h^k & := P^k[(H(\bar{x}, \bar{\lambda}) - M^k) d^k]. \end{array}$$

The condition will be

$$h^k = o(d^k). \tag{5.7}$$

As a particular case of our results we will recover the characterization of Boggs, Tolle and Wang [3] concerning equality constrained nonlinear programming problems, and we will be able to extend the characterization to variational inequalities satisfying the assumption

$$d^t H(\bar{x}, \bar{\lambda}) d > 0 \quad \text{for all } d \text{ in } C - \{0\}. \tag{5.8}$$

All our results, however will need the following qualification hypothesis (equivalent to the linear independence of the gradient of active constraints)

$$g'_{\bar{I}}(\bar{x}) \quad \text{surjective.} \tag{5.9}$$

On the other hand we do not need any strict complementary hypothesis.

Theorem 5.2 *Let $\{(x^k, \lambda^k)\}$ be computed by Algorithm 3 converge towards $(\bar{x}, \bar{\lambda})$, semi-stable solution of (5.2) satisfying (5.9). Then :*

(i) *Condition (5.7) is sufficient for superlinear convergence when E^k is defined as*

$$E_1^k := \{d \in \ker g'_{I^*}(x^k); g'_i(x^k) d \geq 0, \forall i \in I^0 \quad \textit{such that } g'_i(x^k) + g_i(x^k) d^k = 0\}$$

(ii) *Condition (5.7) is necessary, and also sufficient for superlinear convergence if in addition (5.8) holds, when E^k is defined as*

$$E_2^k := \{d \in \ker g'_{I^\bullet}(x^k); g'_{I^0}(\bar{x})d \leq 0\}.$$

Remark 5.1 *If the strict complementarity hypothesis holds, i.e. $I^0 = \emptyset$, then $E_1^k = E_2^k = \ker g'_{I^\bullet}(x^k)$ and, with this choice of E^k, condition (5.7) is necessary and sufficient for superlinear convergence of $\{x^k\}$ (under the hypotheses of Theorem 5.2 but without hypothesis (5.8)).*

Proof of Theorem 5.2

a) Preliminaries. Writing the Kuhn-Tucker conditions for the projection problem defining h^k we get the existence of $\eta^k \in \mathbf{R}^p$ satisfying

$$h^k - [H(\bar{x}, \bar{\lambda}) - M^k]d^k + g'(x^k)^*\eta^k = 0 \tag{5.10}$$

and

$$\begin{aligned}
\eta_i^k &= 0 \quad \text{if } i \in I - \bar{I}, \\
\eta_{I^0}^k &\leq 0, \eta_i^k = 0 \quad \text{for all } i \text{ in } I^0 \text{ such that } g_i(x^k) + g_i'(x^k)d^k > 0, \text{ if } E^k = E_1^k, \\
\eta_{I^0}^k &\geq 0 \quad \text{if } E^k = E_2^k.
\end{aligned}$$

Substracting the first relation defining the Newton step from (5.10) we get

$$h^k - F(x^k) - H(\bar{x}, \bar{\lambda})d^k + g'(x^k)^*(\eta^k - \lambda^{k+1}) = 0. \tag{5.11}$$

Expanding $F(x^k)$ up to the first order and taking (5.2) in account we have

$$\begin{aligned}
-F(x^k) &= -F(\bar{x}) - F'(\bar{x})(x^k - \bar{x}) + o(x^k - \bar{x}), \\
&= g'(\bar{x})^*\bar{\lambda} - F'(\bar{x})(x^k - \bar{x}) + o(x^k - \bar{x}), \\
&= g'(x^k)^*\bar{\lambda} - H(\bar{x}, \bar{\lambda})(x^k - \bar{x}) + o(x^k - \bar{x}),
\end{aligned}$$

hence with (5.11) :

$$h^k - H(\bar{x}, \bar{\lambda})(x^k + d^k - \bar{x}) + g'(x^k)^*(\bar{\lambda} + \eta^k - \lambda^{k+1}) = o(x^k - \bar{x}). \tag{5.12}$$

Let us define

$$\delta^k := \|h^k\| + \|x^k - \bar{x}\| + \|d^k\|.$$

Then $(\delta^k)^{-1}(h^k, x^k + d^k - \bar{x}, \lambda^{k+1} - \eta^k - \bar{\lambda})$ is bounded, the boundedness of the third term being a consequence of (5.9) and (5.12). Let (h, z, ζ) be a limit-point of this sequence i.e. a limit for a subsequence $k \in S \subset \mathbf{N}$. Then $\zeta_i = 0$ if $i \in I - \bar{I}$ and from (5.12) we deduce that

$$h - H(\bar{x}, \bar{\lambda})z - g'(\bar{x})^*\zeta = 0. \tag{5.13}$$

Also, expanding g as follows :

$$g(x^k) + g'(x^k)d^k = g(\bar{x}) + g'(\bar{x})(x^k + d^k - \bar{x}) + o(x^k - \bar{x}),$$

we deduce from the fact that d^k is a Newton step associated to a multiplier λ^{k+1} that

$$\left\{\begin{array}{ll} g'(\bar{x})z \overset{I}{\ll} 0, & \\ g_i'(\bar{x})z = 0, & \text{when } \lambda_i^{k+1} > 0 \text{ for a subsequence, } i \in I^0 \\ & \text{(a fortiori when } \bar{\lambda}_i > 0). \end{array}\right. \tag{5.14}$$

The above relations imply

$$z \in C. \tag{5.15}$$

We also have from the definition of h^k :

$$g'_{I^*}(\bar{x})h = 0, \tag{5.16}$$

$$h \in C \quad \text{when } E^k = E_2^k. \tag{5.17}$$

b) Proof of case i). If $h^k = o(d^k)$, then a fortiori $h = 0$. With (5.13) we deduce that

$$H(\bar{x}, \bar{\lambda})z + g'(\bar{x})^*\zeta = 0. \tag{5.18}$$

If $i \in I^0$ is such that $g_i'(\bar{x})z < 0$, then $g(x^k) + g'(x^k)d^k < 0$ and $\eta_i^k = 0$ and $\lambda_i^{k+1} = 0$ for k in S large enough, hence $\zeta_i = 0$. This implies

$$\zeta_i g_i'(\bar{x})z = 0 \quad \text{for all } i \text{ in } I^0.$$

But this with the fact that $\zeta_{I^0} \geq 0$ (due to $\eta_{I^0}^k \leq 0$), (5.15), (5.18) and (5.6) imply $z = 0$, i.e.

$$x^{k+1} - \bar{x} = o(\|h^k\| + \|x^k - \bar{x}\| + \|d^k\|).$$

This and (5.7) imply

$$x^{k+1} - \bar{x} = o(\|x^k - \bar{x}\| + \|x^{k+1} - x^k\|) = o(\|x^k - \bar{x}\| + \|x^{k+1} - \bar{x}\|)$$

which implies $x^{k+1} - \bar{x} = o(x^k - \bar{x})$, i.e. x^k converges superlinearly.

c) Proof of case ii). If x^k converges superlinearly then $z = 0$. Computing the scalar product of (5.13) by h we get

$$\|h\|^2 = \langle \zeta, g'(\bar{x})h \rangle.$$

Using (5.17), the positivity of $\lambda_{I^0}^{k+1}$ and the complementarity condition $\eta_i^k g_i'(x^k)h^k = 0$, for i in I^0, we deduce that the right hand side of the above relation is non positive ; hence $h = 0$, i.e.

$$h^k = o(\|h^k\| + \|x^k - \bar{x}\| + \|d^k\|)$$

which implies $h^k = o(\|x^k - \bar{x}\| + \|d^k\|)$. However the superlinear convergence of $\{x^k\}$ implies $\|d^k\|/\|x^k - \bar{x}\| \to 1$, hence (5.7) holds.

We now prove that if (5.7) and (5.8) hold, $\{x^k\}$ converges superlinearly. As (5.7) implies $h = 0$, computing the scalar product of (5.13) with z we get

$$\langle z, H(\bar{x}, \bar{\lambda})z \rangle + \langle \zeta, g'(\bar{x})z \rangle = 0. \tag{5.19}$$

As $z \in C, g_i'(\bar{x})z = 0$ if $\lambda_i^{k+1} \neq 0$, and $\eta_{I^0}^k \geq 0$, the second term of (5.19) is non negative. This and (5.8) imply that $z = 0$ i.e.

$$x^{k+1} - \bar{x} = o(\|h^k\| + \|x^k - \bar{x}\| + \|d^k\|).$$

Using (5.7) and the relation $\|d^k\| \leq \|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\|$ we deduce that

$$x^{k+1} - \bar{x} = o(\|x^k - \bar{x}\| + \|x^{k+1} - \bar{x}\|).$$

which implies the superlinear convergence of $\{x^k\}$. □

With the help of Theorem 5.2 we may obtain the superlinear convergence of $\{x^k\}$ when M^k is updated using ideas of quasi-Newton algorithms. We define the quasi-Newton equation (for M^{k+1}) as follows

$$M(x^{k+1} - x^k) = F(x^{k+1}) - F(x^k) + \left[g'(x^{k+1}) - g'(x^k)\right]^* \lambda^{k+1}. \tag{5.20}$$

We assume that there exists a closed convex subset $\mathcal{K}$ of the space of $n \times n$ matrices such that

$$H(x, \lambda) \in \mathcal{K}, \ \forall (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^p, \tag{5.21}$$

and we choose M^{k+1} solution of

$$\min \|M - M^k\|_{\sharp}; \ M \in \mathcal{K}; \ M \quad \text{satisfies (5.20)}, \tag{5.22}$$

where as before $\|\cdot\|_{\sharp}$ is a norm associated to a scalar product.

Lemma 5.1 *Under the hypotheses of Theorem 5.2, if M^{k+1} satisfies (5.20) and*

$$(M^{k+1} - M^k)(x^{k+1} - x^k) = o(x^{k+1} - x^k), \tag{5.23}$$

then $\{x^k\}$ converges superlinearly.

Proof As M^{k+1} satisfies (5.20) we have

$$\begin{aligned} M^{k+1}(x^{k+1} - x^k) &= F(x^{k+1}) - F(x^k) + (g'(x^{k+1}) - g'(x^k))^*\bar{\lambda} + (g'(x^{k+1}) - g'(x^k))^*(\lambda^{k+1} - \bar{\lambda}) \\ &= H(\bar{x}, \bar{\lambda})(x^{k+1} - x^k) + o(x^{k+1} - x^k). \end{aligned}$$

Hence if M^{k+1} satisfies (5.20), and (5.23) holds, then

$$(H(\bar{x}, \bar{\lambda}) - M^k)(x^{k+1} - x^k) = o(x^{k+1} - x^k).$$

This and Theorem 5.2 (case i) imply the conclusion. □

Theorem 5.3 *Let $\{(x^k, \lambda^k)\}$ be computed by Algorithm 3 converge towards $(\bar{x}, \bar{\lambda})$, semi-stable solution of (5.1) satisfying (5.9). If (5.20)-(5.22) hold and*

$$\sum_k \left[\|x^k - \bar{x}\| + \|\lambda^k - \bar{\lambda}\|\right] < \infty, \tag{5.24}$$

then $\{x^k\}$ converges superlinearly.

Proof Define

$$S^k := \{M \in \mathcal{K}; M \quad \text{satisfies (5.20)}\},$$

$$A^k := \int_0^1 H(x^k + \sigma(x^{k+1} - x^k), \lambda^{k+1}) d\sigma.$$

Then A^k is an element of S^k and for some $c_1 > 0$

$$\|A^k - H(\bar{x}, \bar{\lambda})\|_\sharp \leq c_1 \nu^k, \tag{5.25}$$

with here

$$\nu^k := \|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\| + \|\lambda^{k+1} - \bar{\lambda}\|.$$

As M^{k+1} is the projection of M^k onto S^k, we have

$$\|M^{k+1} - M^k\|_\sharp^2 + \|M^{k+1} - A^k\|_\sharp^2 \leq \|M^k - A^k\|_\sharp^2, \tag{5.26}$$

hence with (5.25)

$$\|M^{k+1} - H(\bar{x}, \bar{\lambda})\|_\sharp \leq \|M^k - H(\bar{x}, \bar{\lambda})\|_\sharp + 2c_1\nu^k.$$

This and (5.24) imply that $\|M^k - H(\bar{x}, \bar{\lambda})\|_\sharp$ converges ; as $A^k \to H(\bar{x}, \bar{\lambda})$, $\|M^{k+1} - A^k\|_\sharp$ and $\|M^k - A^k\|_\sharp$ do converge to the same limit, and with (5.25) this implies $M^{k+1} - M^k \to 0$. The conclusion is then a consequence of Lemma 5.1. $\square$

6 Application to nonlinear programming

In this section we particularize some of our results to nonlinear programming problem, and we will see that it allows to get some improvements with respect to known results. By nonlinear programming problem we mean

$$\min f(x); g(x) << 0 \tag{6.1}$$

where f is a mapping $\mathbb{R}^n \to \mathbb{R}$, and g as well as the relation "$<<$" are as in section 5. Let us recall some well-known facts of optimization theory (see e.g. Fletcher [9]). To problem (6.1) is associated the first-order optimality system

$$\begin{cases} \nabla f(x) + g'(x)^* \lambda = 0, \\ g(x) << 0, \ \lambda_I \geq 0, \ \lambda^t g(x) = 0, \end{cases} \tag{6.2}$$

which is formally equivalent to (5.1) if we define $F(x) := \nabla f(x)$. In this case the mapping $H(x, \lambda)$ can be interpretated as the Hessian with respect to x of the Lagrangian $L(x, \lambda) := f(x) + \lambda^t g(x)$. We will say that λ is a Lagrange multiplier associated to x if (x, λ) satisfies (6.2). We recall the results involving second-order conditions involving one multiplier.

Proposition 6.1 *(see e.g. Ben-Tal [1])*

(i) *(Second-order necessary condition). Let x be a local solution of (6.1) to which is associated a unique multiplier λ. Then $d^t H(x, \lambda) d \geq 0$ for all critical direction d.*

(ii) *(Second-order sufficiency condition). Let (x, λ) satisfying (6.2) be such that $d^t H(x, \lambda) d > 0$ for all non zero critical direction d. Then x is a local solution of (6.1).*

We now make the link between semi-stability and the second-order sufficiency conditions.

Proposition 6.2 *Let (x, λ) be an isolated solution of (6.2) such that x is a local solution of (6.1). Then (x, λ) is semi-stable iff it satisfies the second-order sufficiency condition.*

Proof Characterization (3.2) of semi-stability applied to the variational inequality in form (5.1), and using (5.4), gives

$$\langle d, H(x,\lambda)d\rangle \quad > \quad 0 \text{ for all } (d,\mu) \neq 0 \text{ solution of}$$

$$\begin{aligned} H(x,\lambda)d + g'(x)^*\mu &= 0, \\ g(x)^t(\mu - \lambda) &= 0, \\ g(x) + g'(x)d &\in N(\lambda), \end{aligned}$$

the last relation expressing the fact that a multiple of d is critical. Hence the second-order sufficiency optimality condition implies semi-stability. Conversely, let us assume that the second-order sufficiency condition does not hold. By Proposition 6.1 there exists a critical direction $\bar{d} \neq 0$ with $\bar{d}^t H(x,\lambda)\bar{d} = 0$, and $\bar{d}$ is a solution of the quadratic homogeneous problem

$$\min d^t H(x,\lambda)d; \ d \in C.$$

As

$$C = \{d; g'(x)d << 0; g_i'(\bar{x})d = 0 \quad \text{if } \lambda_i > 0, \forall i \in I\},$$

to $\bar{d}$ is associated a multiplier μ such that $(\bar{d}, \mu)$ satisfies (5.6). By Proposition 5.1 this contradicts semi-stability. □

Proposition 6.3 *Let (x, λ) be a semi-stable solution of (6.2) such that x is a local solution of (6.1). Then (x, λ) is hemi-stable.*

Proof Semi-stability implies the uniqueness of the multiplier, hence also the hypothesis of Mangasarian and Fromovitz. By Proposition 6.2 the second-order sufficiency condition also holds for problem (6.1) at (x, λ). It is easy to check that the problem

$$\min_d \nabla f(x)^t d + \frac{1}{2} d^t H(x,\lambda)d; \ g(x) + g'(x)d << 0$$

has $\bar{d} = 0$ as an isolated local solution associated to the unique multiplier λ. Hence if we make a small perturbations in the data of this problem there exists a local solution whose distance to $\bar{d} = 0$ is of the order of the perturbation (see e.g. Bonnans [5]). Hence hemi-stability holds. □

From Theorem 2.2 and Proposition 6.3 we deduce

Theorem 6.1 *Assume that x is a local solution of (6.1), λ is the unique Lagrange multiplier associated to x, and the weak second-order sufficiency condition holds. Then there exists $\varepsilon > 0$ such that if $\|x^0 - x\| + \|\lambda^0 - \lambda\| < \varepsilon$, and (x^{k+1}, λ^{k+1}) is chosen so that $\|x^{k+1} - x^k\| + \|\lambda^{k+1} - \lambda^k\| < \varepsilon$, then Algorithm 3 with $M^k = H(x^k, \lambda^k)$, i.e. Newton's method, is well defined and converges at a quadratic rate to (x, λ).*

Remark 6.1 *That Newton's method converges at a quadratic rate when the starting point is close to a solution* (x, λ) *of (6.2), assuming* x *is a local solution of (6.1), the gradient of active constraints linearly independent, and strict complementarity, is well known. Recently the author [4] relaxed the strict complementarity hypothesis. Here we improve the result of [4] by assuming that the multiplier is unique instead of the linear independence of the gradients of active constraints.*

We now apply the results of section 5 on the superlinear convergence of $\{x^k\}$ only. From Theorem 5.2 and the fact that condition (5.8) coincides with the second-order sufficient condition we get

Theorem 6.2 *Let* $\bar{x}$ *be a local solution of (6.1) such that the gradient of active constraints are linearly independent,* $\bar{\lambda}$ *be a multiplier associated to* $\bar{x}$ *and the second-order sufficient condition holds. Then if* (x^k, λ^k) *computed by successive quadratic programming converge to* $(\bar{x}, \bar{\lambda})$*, then* $\{x^k\}$ *converges superlinearly iff*

$$P^k[(H(\bar{x}, \bar{\lambda}) - M^k)d^k] = o(d^k),$$

with P^k *orthogonal projection on the set* E_2^k *defined in Theorem 5.2.*

Remark 6.2 *If no inequality constraint is present, Theorem 6.2 reduces to a theorem of Boggs, Tolle and Wang [3]. Some necessary or sufficient conditions (but not the characterization given here) for problems with equalities and inequalities, without strict complementarity have been given by the author in [4].*

References

[1] A. Ben-Tal : *Second-order and related extremality conditions in nonlinear programming.* J. Optimiz. Theory Appl. 21 (1980), 143-165.

[2] D.P. Bertsekas : *Constrained optimization and Lagrange multipliers methods.* Academic Press, New York (1982).

[3] P.T. Boggs, J.W. Tolle, P. Wang : *On the local convergence of quasi-Newton methods for constrained optimization.* SIAM J. Control Optimiz. 20 (1982), 161-171.

[4] J.F. Bonnans : *Local study of Newton type algorithms for constrained problems.* Proc. 5th Franco-German conf. on Optimization, S. Dolecki ed., Lectures notes in Math. 1405, Springer-Verlag (1989), 13-24.

[5] J.F. Bonnans : *On the stability of solutions in nonlinear programming.* Optimization 21 (1990).

[6] J.F. Bonnans : *Méthodes newtoniennes.* SMAI series on "Mathématiques et Applications", Ellipses, Paris (book to appear).

[7] H. Brézis : *Opérateurs maximaux monotones.* North-Holland, Amsterdam, 1973.

[8] J.E. Dennis, J.J. Moré : *Quasi-Newton methods, motivation and theory.* SIAM Review 19 (1977), 46-89.

[9] R. Fletcher : *Practical methods of optimization.* (2nd edition) Wiley, Chichester (1987).

[10] D. Gabay : *Application de la méthode des multiplicateurs aux inéquations variationnelles.* In "Méthodes de lagrangien augmenté", M. Fortin and R. Glowinski ed., Dunod, Paris, 1982, 279-308.

[11] S.M. Grzegòrski : *Orthogonal projections on convex sets for Newton-like methods*, SIAM J. Numer. Anal. 22 (1985), 1208-1219.

[12] N.H. Josephy : *Newton's method for generalized equations.* Technical Summary Report No 1965, Mathematics Research Center, University of Wisconsin-Madison (1979).

[13] N.H. Josephy : *Quasi-Newton methods for generalized equations.* Summary Report No 1966, Mathematics Research Center, University of Wisconsin-Madison (1979).

[14] J.L. Lións, G. Stampacchia : *Variational inequalities.* Comm. Pure Appl. Math. 20 (1967), 493-519.

[15] O.L. Mangasarian : *Locally unique solutions of quadratic programs, linear and nonlinear complementarity problems.* Math. Programming 19 (1980), 200-212.

[16] A. Reinoza : *The strong positivity condition.* Math. Oper. Res. 10 (1985), 54-62.

[17] S.M. Robinson : *Strongly regular generalized equations.* Math. Oper. Res. 5 (1980), 43-62.

[18] S.M. Robinson : *Generalized equations.* In "Mathematical Programming, the state of the Art", A. Bachem, M. Gröstchel, B. Korte ed., Springer-Verlag (1983), 346-367.

[19] S.M. Robinson : *Newton's method for a class of nonsmooth functions.* To appear.

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique

ISSN 0249-6399